J Chron Dis Vol. 39, No. 7, pp. 513–519, 1986
Printed in Great Britain

0021-9681/86 $3.00 + 0.00
Pergamon Journals Ltd

SELECTING SUBPOPULATIONS FOR INTERVENTION

SHOLOM WACHOLDER[1] and CLARICE R. WEINBERG[2]

[1]McGill University, Department of Epidemiology and Biostatistics, 1020 Pine Avenue West, Montreal, Quebec, Canada H3A 1A2 and
[2]National Institute of Environmental Health Sciences, Biometry and Risk Assessment Program, P.O.B. 12233, Research Triangle Park, NC 27709, U.S.A.

(*Received in revised form* 2 *January* 1986)

Abstract—The effectiveness of a program designed to eliminate exposure to a risk factor may be enhanced by selecting specific subpopulations for the intervention effort. A second factor is called an intervention effect modifier of the first factor if the effect of the intervention on the first factor depends on the distribution of the second factor in the population chosen for the intervention. If an intervention effect modifier exists, a program's effectiveness may be enhanced by choosing a subpopulation for intervention on the basis of the distribution of that factor. We discuss the conditions in which intervention effect modification can occur when there is a third factor which influences risk. Additivity of incidence rates, the criterion when only two factors are involved, is neither necessary nor sufficient to rule out intervention effect modification in this situation. We also discuss the cost effectiveness of intervention strategies when the cost per successful intervention varies.

INTRODUCTION

IT IS NOW generally agreed that additive rather than multiplicative risk models should be used to assess public health effects. For example, Kleinbaum *et al.* assert that "departures from additivity of risk differences should be the focus whenever public health issues are involved" ([1] p. 412). We show in this paper that in some realistic situations reliance on additivity of risk differences can be misleading in the assessment of public health effects because of confounding by an extraneous factor.

Planners of programs to eliminate exposure to harmful agents must consider which subpopulations to target for the intervention effort. If an extraneous factor influences the risk reduction obtained by intervention, the biggest public health impact will be obtained by selecting those individuals in whom the greatest reduction in risk is expected. For example, convincing 10,000 men to quit smoking will prevent more deaths than would be prevented by convincing a combination of 10,000 men and women to quit, if the reduction in mortality for males who quit smoking is greater than for females. We define a factor B to be an *effect modifier of the intervention* on exposure A if the effect of removing exposure to A in a population selected for intervention depends on the distribution of exposure to B in that population. So, if B is not an intervention effect modifier of A, B can be ignored when planning interventions aimed at elimination of exposure to A.

We have the following goals in this paper:

(1) To show that there are two distinct concepts of public health interaction in the literature, to disentangle them, and to clarify terminology;

(2) To show by example that intervention effect modification, just as direct effects, can be confounded by a third factor;

(3) To derive criteria for intervention effect modification in a simple situation;

(4) To show how cost should be accounted for in planning public health interventions.

Material may be protected by copyright law (Title 17, U.S. Code)

We begin with several assumptions, made implicitly by others [2–5]. We presume that if the intervention succeeds in an individual, that individual will no longer be exposed to A thereby reducing his risk to background levels, but if the intervention fails, the individual's exposure to A is unchanged. We also assume that intervention on one factor does not affect exposure to other factors.

INCIDENCE ADDITIVITY FOR TWO FACTORS

Rothman *et al.* [2] discuss the problem of which subgroup to select for intervention. They consider dichotomous factors A and B to not exhibit *public health interaction* when the expected number of incident cases in a population does not depend on the joint distribution of A and B for fixed marginal distributions of A and B; i.e. when the number of new cases does not depend on the number of individuals with both exposures, but only on the respective numbers of individuals exposed to each [2]. If this condition does not hold, there is *public health interaction* between A and B. When only two risk factors are considered they show that this definition of public health interaction is equivalent to a departure from additivity of incidence rates.

Assume that the measure of interest is incidence per year of a particular disease and that incident cases of this disease occur as a random process which is constant over time and depends only on two factors, A and B. Let I_{ij} be the incidence rate at level i of A and level j of B, with level 0 of A representing no exposure or background exposure and 1 representing exposure, and let N_{1j} denote the number of individuals at level 1 of A and j of B who are to receive intervention. The dot subscript will be used to indicate summation over that subscript. The reduction in incidence rate achieved by changing the level of A from 1 to 0 in a population at level j of B is then $I_{1j} - I_{0j}$. So the effect of the intervention will be a reduction in incidence of

$$\sum_j N_{1j}(I_{1j} - I_{0j})$$

which equals

$$(I_{1j} - I_{0j}) \sum_j N_{1j}$$

whenever the differences $I_{1j} - I_{0j}$ are equal for all j. In other words, the total reduction is independent of the distribution of B in those receiving the intervention when the incidence difference is the same for all j. Thus, as noted by Rothman *et al.* [2], there will be no intervention effect modification when the *additive* or *no interaction model of incidences*,

$$I_{1j} - I_{0j} = I_{10} - I_{00} \tag{1}$$

holds for all j. When equation (1) holds, the reduction in incidence achieved by intervention in a fixed number of people exposed to A is the same, whether the intervention is given preferentially to those at a particular level of B or independently of B. Thus equation (1) is the criterion for both absence of public health interaction and absence of intervention effect modification. Equation (1) also represents the condition of no *effect modification*, in its usual sense, of B on A on the risk difference scale.

The hypothetical data in Table 1A is an example of lack of intervention effect modification and of lack of public health interaction. The Table contains hypothetical incidence rates at each of the four combinations of levels of dichotomous factors A and B. In the rows representing individuals exposed to A, column 3 represents the incidence rate in those unexposed to A, at the same level of B. Column 4 is the expected reduction in incidence if 1000 people eliminate their exposure to A. Since the reduction in incidence per individual receiving intervention is 24/1000 at both levels of B, intervention is equally efficacious regardless of the distribution of B among those receiving it.

Material may be protected by copyright law (Title 17, U.S. Code)

TABLE 1A. APPARENT ABSENCE OF INTERVENTION EFFECT MODIFICATION IN A POOLED TABLE

A B	Population	Incidence per 1000/yr	Presumed incidence after removal of A	Presumed incidence difference
0 0	3000	8	—	—
0 1	9000	16	—	—
1 0	9000	32	8	24
1 1	27000	40	16	24

TABLE 1B. A POSSIBLE BREAKDOWN OF TABLE 1A[1]

A B C	Population	Incidence per 1000/yr	Incidence after change in A	Incidence difference
0 0 0	2000	6	—	—
0 0 1	1000	12	—	—
0 1 0	6000	12	—	—
0 1 1	3000	24	—	—
1 0 0	6000	24	6	18
1 0 1	3000	48	12	36
1 1 0	18000	30	12	18
1 1 1	9000	60	24	36

[1]This tabulation is consistent with Table 1A. $I_{ijk} = (I_{i00} + I_{0j0} - I_{000}) * I_{00k}/I_{000}$ so, on the additive scale, A and B interact with C, but not with each other. Exposures to A, B, and C occur independently. The effect of intervention in those unexposed to B is a reduction of $(6*18 + 3*36)/9 = 24$ cases per year. The effect of intervention in those exposed to B is a reduction of $(18*18 + 9*36)/27 = 24$ cases per year.

TABLE 1C. ANOTHER POSSIBLE BREAKDOWN OF TABLE 1A[1]

A B C	Population	Incidence per 1000/yr	Incidence after change in A	Incidence difference
0 0 0	2000	8	—	—
0 0 1	1000	8	—	—
0 1 0	6000	12	—	—
0 1 1	3000	24	—	—
1 0 0	3000	48	8	40
1 0 1	6000	24	8	16
1 1 0	12000	40	12	28
1 1 1	15000	40	24	16

[1]This tabulation is also consistent with Table 1A. Exposures to A, B, and C do not occur independently. The effect of intervention in those unexposed to B is a reduction of $(3*40 + 6*16)/9 = 24$ cases per year. The effect of intervention in those exposed to B is a reduction of $((12*28) + (15*16))/27 = 21.3 < 24$ cases per year.

Walter [4, 5] addresses a third kind of "public health interaction," which we call *intervention non-additivity*: whether the sum of the effects of two separate interventions, one eliminating A and the other eliminating B, does not equal the effect of eliminating both together. For example, a campaign to reduce highway mortality might try to reduce drunk-driving, to reduce the percentage of drivers who do not use seatbelts, or to simultaneously reduce both. The expected reduction in mortality which will be achieved by eliminating exposure to A is

$$\mathrm{Red}_A = \sum_j N_{1j}(I_{1j} - I_{0j}).$$

Similar expressions are obtained for Red_B and $\mathrm{Red}_{A,B}$. Walter [5] showed that $\mathrm{Red}_A + \mathrm{Red}_B = \mathrm{Red}_{A,B}$ if and only if

$$\sum_j N_{1j}(I_{1j} - I_{0j} - I_{10} + I_{00}) = 0 \tag{2}$$

That is, additivity is required on average across B, but not necessarily within each exposure category. Additivity of incidence rates is a sufficient condition for this equality, but is not generally necessary except when A and B are both dichotomous.

While intervention additivity, lack of intervention effect modification, and incidence rate additivity are equivalent for two dichotomous risk factors, they are not equivalent in more

Material may be protected by copyright law (Title 17, U.S. Code)

TABLE 2. INTERVENTION ADDITIVITY WITHOUT INCIDENCE ADDITIVITY

			Incidence reduction after changing exposure to level 0 of		
A B	Population	Incidence per 1000/yr	A	B	A and B
0 0	1000	10	0	0	0
0 1	1000	14	0	4	4
0 2	1000	18	0	8	8
1 0	1000	18	8	0	8
1 1	1000	21	7	7	11
1 2	1000	23	5	5	13
Cases prevented			20	24	44

complex situations. Table 2 is an example. Removing exposure to A in individuals at level 0 of B reduces the incidence rate by 8/1000 which is greater than the reductions for $B = 1$ and $B = 2$. Thus B is an intervention effect modifier of A and the incidence rates are not additive. However, a program which intervenes on A alone in the entire population would reduce the number of cases by 20, on B alone by 24, and on A and B simultaneously by $44 = 20 + 24$. Thus we have an example of intervention additivity together with intervention effect modification.

When there are two risk factors which are not dichotomous, intervention effect modification and Rothman's public health interaction (nonadditivity of rates) are equivalent. Additivity of rates is sufficient but not necessary for intervention additivity.

EFFECT OF MORE THAN TWO RISK FACTORS

Intervention effect modification is similar to classical effect modification [1] where difference in incidence rates is the measure of association. Effect modification is simply heterogeneity of a measure across strata; but the effect of intervention on an exposure may vary with the stratum, even when there is no effect modification in the incidence difference. The reason for the discrepancy is that a third factor may confound the interaction between the stratum variable and the exposure. If exposure to A is eliminated in those at level j of B, the resulting incidence rate may not be I_{0j} since A may be associated with a third factor related to risk but unchanged by the intervention. Reliance on additive models, as in Rothman's definition of public health interaction, is inadequate for intervention planners when there is a third factor involved.

Let C represent all other factors which influence the incidence of disease or may act as effect modifiers for A or B. We first assume that C status will be known in individuals who will receive intervention. With k denoting the level of C, elimination of exposure to A will have the same impact regardless of the joint distribution of B and C, that is, B and C will not to be joint intervention effect modifiers, when

$$I_{1jk} - I_{0jk} = I_{100} - I_{000} \tag{3}$$

for all j and k. This is an incidence model with main effects for A, B, and C and only the B–C interaction. Additivity of the incidences in A, B, and C satisfies this condition. The weaker assumption that A has no interaction with B or with C also implies that B does not modify the intervention, but no constraint on the B–C interaction needs to be imposed. Thus an additive model in A, B, and C is a sufficient but not a necessary condition for B and C not to be joint intervention effect modifiers. Table 1B represents a possible breakdown of Table 1A. In Table 1B, B and C are joint intervention effect modifiers: note that at each level of B, targetting individuals with $C = 1$ will reduce incidence by twice as much as with $C = 0$.

Often there will be a major risk factor represented by C which is unknown or will be too risky or expensive to ascertain for each individual considered for the intervention program. Reliance on equation (1) for determination of intervention effect modification by B can be misleading in such a case. Assume that selection for intervention within each

rial may be protected by copyright law (Title 17, U.S. Code)

level of B is independent of C. Since the reduction in incidence for a person with $B = j$ and $C = k$ who changes from $A = 1$ to $A = 0$ will be $I_{1jk} - I_{0jk}$, the number of cases prevented per person receiving intervention with $B = j$ is

$$\sum_k \frac{N_{1jk}(I_{1jk} - I_{0jk})}{N_{1j.}}.$$

If there is no intervention effect modification, then these are equal for all j; i.e.

$$\sum_k \frac{N_{1jk}(I_{1jk} - I_{0jk})}{N_{1j.}} = \sum_k \frac{N_{10k}(I_{10k} - I_{00k})}{N_{10.}}. \tag{4}$$

for all j.

When will B not modify the effect of intervention on A? Inspection of equation (4) reveals that absence of A–B interaction at each level of C is neither sufficient nor necessary: the weights given to the incidence differences on the left and right hand sides of the equations differ. However, absence of an A–B interaction at each level of C *is* sufficient when C is independent of B in those exposed to A; the A–B interaction cannot be confounded by C when C is independent of B in those exposed to A. But when C is not independent of B, C can confound the A–B interaction. So additivity in A and B is neither sufficient nor necessary for lack of intervention effect modification of A by B.

We can rewrite equation (1) in terms of the I_{ijk} as

$$\sum_k \frac{N_{1jk} I_{1jk}}{N_{1j.}} - \sum_k \frac{N_{0jk} I_{0jk}}{N_{0j.}} = \sum_k \frac{N_{10k} I_{10k}}{N_{10.}} - \sum_k \frac{N_{00k} I_{00k}}{N_{00.}}. \tag{5}$$

for all j, by replacing each I_{ij} with appropriate weighted averages of the I_{ijk}. The difference between equations (4) and (5) is that the incidence rate after intervention in equation (4) is based on the distribution of C in those exposed to A at level j of B who are to receive intervention, while the corresponding term in equation (5) is weighted by the irrelevant distribution of C in those unexposed to A before the intervention period.

Inference about intervention effect modification based on equation (1) is appropriate only if equations (4) and (5) are equivalent. When conclusions based on equation (5) differ from those which would be drawn from equation (4), we say that factor C is a confounder of the intervention effect modification. In general, factor C can confound the intervention effect modification by B on A if, at some level of B, C is related to both excess disease incidence and to A. In the absence of either of these conditions, additivity in risk across A and B is sufficient to ensure that there will be no intervention effect modification by B on A due to C.

For the population in Table 1A, the incidence rates are additive in A and B and equation (1) is satisfied. So intervention planners who relied on the additive model would anticipate that B would not modify the effect of intervention on A. Tables 1B and C represent two further hypothetical breakdowns of the population in Tables 1A, but include information about C which might not be available to the planners of the intervention. In the population represented in 1B, the effect of intervention on A in 1000 individuals unexposed to B is a reduction of $(6/9)*18 + (3/9)*36 = 24$ cases, if selection of individuals for intervention is made independently of C. Similarly, the effect among those exposed to B is a reduction of $(18/27)*18 + (9/27)*36 = 24$ cases. But using the alternative population represented in Table 1C, which is also consistent with Table 1A, intervention in 1000 individuals unexposed to B would result in a reduction in incidence of 24, while a reduction of only 21.3 would be achieved by selecting those exposed to B. So, without more information or additional assumptions, program planners cannot be sure of any conclusion on intervention effect modification based on the rates in Table 1A, because C may confound the intervention effect modification of by B on A, as in Table 1C. This example demonstrates that additivity of incidences in a pooled table does not ensure absence of intervention effect modification.

Material may be protected by copyright law (Title 17, U.S. Code)

Note that equation (4) holds for both 1B and 1C, but equation (5) holds only for Table 1B. Moreover, even though the incidence rates in 1B do not following an additive model in A, B, and C, factor B is not an intervention effect modifier of A. Modifying the original data in Table 1B so that $N_{1jk} = 9$ for $j = 0, 1$; $k = 0, 1$ produces an example where there is intervention effect modification by A on B but no intervention effect modification by B on A.

The idea discussed here is a generalization of that introduced by Rothman *et al.* [2] where we now allow for possible effects of additional unmeasured factors. Note, however, that intervention effect modification is not symmetric in A and B, since intervention effect modification by B on A can occur without intervention effect modification by A on B. Therefore, words such as interaction and independence which suggest symmetry can be misleading in this context, and new terminology is needed.

Intervention additivity, unlike intervention effect modification, is symmetric in A and B. The condition under which A and B exhibit intervention additivity is

$\mathrm{Red}_A + \mathrm{Red}_B = \mathrm{Red}_{A,B}$; i.e. (when B takes on only two values)

$$\sum_j \sum_k N_{1jk}(I_{1jk} - I_{0jk}) + \sum_i \sum_k N_{i1k}(I_{i1k} - I_{i0k}) = \sum_i \sum_j \sum_k N_{ijk}(I_{ijk} - I_{00k})$$

or

$$\sum_k N_{11k}(I_{11k} - I_{01k} - I_{10k} + I_{00k}) = 0. \tag{6}$$

So, as Walter [5] showed, incidence rate additivity is sufficient but not necessary for equation (6). If C is related to incidence, equation (6) is not equivalent to equations (4) or (5). Note that equation (6), unlike equation (4), is simply a sum of interaction effects weighted by the appropriate number of individuals.

For the population in Table 1B, equations (5) and (6), but not (4), will be satisfied without an additive model if I_{110} is changed from 30 to 36 and I_{111} is changed from 60 to 48. This illustrates that B can modify the effect of intervention on A, even when A and B exhibit intervention additivity.

COST CONSIDERATIONS

Until now, we have not considered the possibility that the cost per successful intervention on A might depend on B. For example, it may be cheaper to vaccinate people in urban areas than rural areas. Furthermore, the proportions of exposed individuals targeted for intervention who actually eliminate their exposure to A may differ in the subpopulations. For example, it may be cheaper to contact men for a stop smoking program, but targeting women might nevertheless be more cost effective if women who are reached are more likely to quit.

Let c_j be the cost per person reached for intervention in the group exposed to A and at level j of B, and let p_j be the proportion of those reached in that exposure group who respond by eliminating exposure to A. Then B does not modify the effect of the intervention on A, in the practical sense that a fixed cost will produce the same benefit regardless of the level of B, when

$$\frac{p_j(I_{1j} - I_{0j})}{c_j} = \frac{p_0(I_{10} - I_{00})}{c_0} \tag{7}$$

for all j. In words, equation (7) says that the cost per case prevented is independent of B. When the cost per successful intervention is independent of B, i.e. when p_j is proportional to c_j, equation (7) reduces to equation (1). Extending cost effectiveness considerations to situations where there is a third factor related to incidence requires complex assumptions about how the three factors jointly affect accessibility and probability of response to the intervention attempt.

Material may be protected by copyright law (Title 17, U.S. Code)

CONCLUSION

Assessment of intervention effect modification from incidence rate models is difficult. Models which are misspecified by ignoring an important risk factor cannot be relied upon. Particularly, exclusive reliance on additive models can lead to spurious conclusions about intervention effect modification when more than two risk factors are involved. When a third factor C related to disease incidence has not been identified, or when assignment to intervention can be based on B but not C, an additive model in A and B does not imply that there is no advantage to targetting people on the basis of their level of B because of the possibility of confounding of the interaction between A and B. also, intervention effect modification by B on A does not imply intervention effect modification by A on B. Finally, if the goal is to make the intervention as cost effective as possible, the cost per successful intervention must be considered.

Intervention non-additivity and intervention effect modification are not the same. The former addresses the joint effects of interventions on two factors while the latter addresses the effect of a single intervention on subgroups defined by an exposure factor which is presumed to be unaffected by the program. Except when there are only two dichotomous risk factors, an additive incidence rate model is a stronger than necessary condition for intervention additivity.

Measures other than reduction in incidence can be used to assess an intervention. Like effect modification, intervention effect modification is measure dependent. Conclusions about modification of intervention effects based on incidence rates may not be applicable when other measures of the effect of an intervention such as reduction in cumulative risk over a period of time, or reduction in lifetime risk, are used.

The argument has been presented in an idealized setting. First, the incidence rates and joint exposure distributions were assumed to be known and to remain constant over time. Moreover, we have assumed that the effect of the intervention is either to leave the exposure unchanged or eliminate it entirely. We have also assumed that the effect of intervention on one risk factor will not effect any other. In practice, this assumption will be violated if, for example, encouraging exercise leads to a healthier diet.

In summary, incidence rate estimates from a model which fits the data well, together with knowledge of the distribution of risk factors in the population in which intervention is planned, can be used to compare the effect of alternative intervention strategies. Because the effects of interventions are usually complicated by unknown factors and unpredictable response patterns, these comparisons need to be tempered by experience and judgment.

Acknowledgements—The authors wish to acknowledge the constructive contribution of Dr Jean-François Boivin to this manuscript.

REFERENCES

1. Kleinbaum DG, Kupper LL, Morgenstern H: **Epidemiologic Research: Principles and Quantitative Methods.** Belmont, CA: Lifetime Learning Publications, 1982
2. Rothman KJ, Greenland S, Walker AM: Concepts of interaction. **Am J Epidemiol** 112: 467–469, 1980
3. Rothman KJ: Occam's razor pares the choice among statistical models. **Am J Epidemiol** 115: 347–349
4. Walter SD: Prevention for multifactorial diseases. **Am J Epidemiol** 112: 409–416
5. Walter SD: Effect of interaction, confounding, and observational error on attributable-risk estimation. **Am J Epidemiol** 117: 598–604

Material may be protected by copyright law (Title 17, U.S. Code)